

Hive

Screen Shots of the analysis

1. Launching an Amazon EMR cluster

The screenshot shows the AWS Management Console interface for an Amazon EMR cluster named 'HiveCluster'. The cluster is in the 'Starting' state. The left sidebar contains navigation links for Amazon EMR, EMR Studio, EMR Serverless, EMR on EC2, Clusters, Notebooks, Git repositories, Security configurations, Block public access, VPC subnets, Events, EMR on EKS, Virtual clusters, Help, and What's new. The main content area displays the cluster's configuration details, including its ID (j-1SG29RU1GHMO), creation date (2023-03-01 08:26 UTC+5:30), and various settings like Release label, Hadoop distribution, Applications, Log URI, EMRFS consistent view, Custom AMI ID, Amazon Linux Release, Availability zone, Subnet ID, Master and Core instance types, Task type, Cluster scaling, and Auto-termination. The 'Summary' tab is selected, showing a progress bar for the cluster's state.

The screenshot shows the AWS Management Console interface for the same Amazon EMR cluster 'HiveCluster', now in the 'Waiting' state. The cluster is ready to run steps. The left sidebar is identical to the previous screenshot. The main content area displays the cluster's configuration details, including its ID (j-1SG29RU1GHMO), creation date (2023-03-01 08:26 UTC+5:30), and various settings like Release label, Hadoop distribution, Applications, Log URI, EMRFS consistent view, Custom AMI ID, Amazon Linux Release, Availability zone, Subnet ID, Master and Core instance types, Task type, Cluster scaling, and Auto-termination. The 'Summary' tab is selected, showing a progress bar for the cluster's state. The 'Application user interfaces' section shows 'YARN timeline server, Tez UI' and 'On-cluster user Not Enabled'. The 'Security and access' section shows 'Key name: vockey', 'EC2 instance profile: EMR_EC2_DefaultRole', 'EMR role: EMR_DefaultRole', and 'Visible to all users: All'.

2. Create an S3 Bucket and Upload data and CSV file to S3 bucket.

The screenshot shows the AWS Management Console for the 'Amazon S3' service. A green notification banner at the top states: "Successfully created bucket 'hiveclusterassignmnet1'". Below this, the 'Buckets' page is displayed. It includes an 'Account snapshot' section with metrics: Total storage (3.0 MB), Object count (394), and Average object size (7.7 KB). The 'Buckets (2)' section shows a table of existing buckets.

Name	AWS Region	Access	Creation date
aws-logs-055500521311-us-east-1	US East (N. Virginia) us-east-1	Objects can be public	January 26, 2023, 18:11:58 (UTC+05:30)
hiveclusterassignmnet1	US East (N. Virginia) us-east-1	Bucket and objects not public	March 1, 2023, 08:28:20 (UTC+05:30)

The screenshot shows the 'Upload: status' page in the AWS Management Console. It displays a summary of the upload process, including the destination (s3://hiveclusterassignmnet1/input/) and the status (Succeeded). The 'Files and folders' section shows a table of the uploaded file.

Name	Folder	Type	Size	Status	Error
Files and folders (1 Total, 99.6 KB)					

3. Connecting to the Hadoop main node by using SSH

[illegible]

4. Running Hive interactively

```
hadoop@ip-172-31-23-86:~$ ssh -i labuser.pem hadoop@ec2-54-235-3-143.compute-1.amazonaws.com
hadoop@ip-172-31-23-86:~$ ssh -i labuser.pem hadoop@ec2-54-235-3-143.compute-1.amazonaws.com
The authenticity of host 'ec2-54-235-3-143.compute-1.amazonaws.com (54.235.3.143)' can't be established.
ED25519 key fingerprint is SHA256:pjA8500bNikuxdq2YwC0xeWtruxX/kstFhhHf8vSg.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-54-235-3-143.compute-1.amazonaws.com' (ED25519) to the list of known hosts.
Last login: Wed Mar 1 03:01:37 2023
┌───┴───┐
┌─┴─┐ Amazon Linux 2 AMI
└─┴─┘
https://aws.amazon.com/amazon-linux-2/
1 package(s) needed for security, out of 8 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEE MWWWWWW MWWWWWW RRRRRRRRRRRRRR
E:::EEEEEEEEEE::E M:::MM M:::MM R:::RRRRRRRRRRRR
EE:::EEEEEEEEEE::E M:::MM M:::MM R:::RRRRRRRRRRRR
E:::E EEEE M:::MM M:::MM RRR::R R:::R
E:::E M:::MM M:::MM R::R R:::R
E:::EEEEEEEEEE M:::MM M:::MM R::RRRRRRRRRRRR
E:::EEEEEEEEEE M:::MM M:::MM R:::RRRRRRRRRRRR
E:::EEEEEEEEEE M:::MM M:::MM R:::RRRRRRRRRRRR
E:::EEEEEEEEEE M:::MM M:::MM R::RRRRRRRRRRRR
E:::E EEEE M:::MM M:::MM R::R R:::R
EE:::EEEEEEEEEE::E M:::MM M:::MM R:::R R:::R
E:::EEEEEEEEEE::E M:::MM M:::MM R:::R R:::R
EEEEEEEEEEEEEEEEEE MWWWWWW MWWWWWW RRRRRRR RRRRRR

hadoop@ip-172-31-23-86 ~$ sudo chown hadoop -R /var/log/hive
hadoop@ip-172-31-23-86 ~$ mkdir /var/log/hive/user/hadoop
hadoop@ip-172-31-23-86 ~$ aws cs ls /
2023-01-26 12:41:58 aws-logs-05560521311-us-east-1
2023-03-01 02:58:20 hiveclusterassigment1
hadoop@ip-172-31-23-86 ~$ hive -d SAMPLES3://aws-tc-largeobjects/CUR-TF-200-ACDENG-1/enr-lab -d DAY=2009-04-13 -d HOUR=08 -d NEXT_DAY=2009-04-13 -d NEXT_HOUR=09 -d OUTPUT=S3://hiveclusterassigment1/output/
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive>
```

5. Creating tables out of source data by using Hive and load data

```
hadoop@ip-172-31-23-86:~$  
hive> CREATE TABLE IF NOT EXISTS delay_flights (  
  > id int,  
  > Year int,  
  > Month int,  
  > DayOfMonth int,  
  > DayOfWeek int,  
  > DepTime string,  
  > CRSDepTime string,  
  > ArrTime string,  
  > CRSArrTime string,  
  > UniqueCarrier string,  
  > FlightNum int,  
  > TailNum string,  
  > ActualElapsedTime string,  
  > CRSElapsedTime string,  
  > AirTime string,  
  > ArrDelay string,  
  > DepDelay string,  
  > Origin string,  
  > Dest string,  
  > Distance int,  
  > TaxiIn int,  
  > TaxiOut int,  
  > Cancelled int,  
  > CancellationCode string,  
  > Diverted int,  
  > CarrierDelay string,  
  > WeatherDelay string,  
  > NASDelay string,  
  > SecurityDelay string,  
  > LateAircraftDelay string)  
  > COMMENT 'delay flights Table'  
  > ROW FORMAT DELIMITED  
  > FIELDS TERMINATED BY ',';  
OK  
Time taken: 3.749 seconds  
hive> LOAD DATA INPATH 's3://hiveclusterassignment1/input/DelayedFlights-updated.csv' INTO TABLE delay_flights;  
FAILED: SemanticException Line 1:17 Invalid path 's3://hiveclusterassignment1/input/DelayedFlights-updated.csv'  
hive> LOAD DATA INPATH 's3://hiveclusterassignment1/input/DelayedFlights-updated.csv' INTO TABLE delay_flights;  
Loading data to table default.delay_flights  
OK  
Time taken: 2.401 seconds  
hive>
```

6. Querying and analyzing the resulting dataset

```
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:686)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:244)
at org.apache.hadoop.util.RunJar.main(RunJar.java:158)
FAILED: ParseException line 1:0 cannot recognize input near 'ELECT' 'Year' ','
hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights where Year is Not NULL GROUP BY Year;
Query ID = hadoop_20230301031605_95199f72-e564-49a7-b3ed-82853987ba89
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)

-----
      VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
map 1 ..... container  SUCCEEDED      1          1          0          0          0          0
reducer 2 ..... container  SUCCEEDED      2          2          0          0          0          0
-----
VERTICES: 02/02 [#####] 100% ELAPSED TIME: 6.39 s
-----
OK
year      _c1
2003      24.557549755575373
2007      19.850007017971283
2009      20.33058554239575
2010      21.89310246015957
2014      43.64459443230066
2015      28.01977637202288
2016      30.453296261292596
2018      28.88346981456985
Time taken: 11.998 seconds, Fetched: 8 row(s)
hive>
```

7. Other results

```
hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year
> ;
Query ID = hadoop_20230301031928_c2fb733d-a57d-4dec-919f-12d30c8dc070
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 7.07 s

PK	year	_c1
0003	24.557549755575373	
0007	19.8500070517971283	
0009	28.33058554239575	
0010	21.89310246015957	
0004	43.64459443230066	
0005	28.01977637202288	
0006	30.453296261292596	
0008	28.88346981456985	

Time taken: 7.974 seconds, Fetched: 8 row(s)

```
hive> SELECT Year, avg((NASDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year
> ;
Query ID = hadoop_20230301032036_37f6f682-0fc8-4c92-a5ca-ad6920b5b170
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.30 s

PK	year	_c1
0003	29.686276314267346	
0007	30.625925917941924	
0009	37.63093330628511	
0010	33.87351363404217	
0004	18.24570061769958	
0005	16.63060805373129	
0006	18.119312329937703	
0008	30.16552562594132	

Time taken: 7.042 seconds, Fetched: 8 row(s)

```
Time taken: 7.042 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year
> ;
Query ID = hadoop_20230301032125_fb1ee526-5c01-49ea-bc0c-d3963ff08afd
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 8.03 s

PK	year	_c1
0003	7.8319479664511205	
0007	4.042975783210207	
0009	0.45316615137982363	
0010	2.9023312955584664	
0004	6.4475279976916555	
0005	5.85069715149616	
0006	4.588604183967953	
0008	3.7254490054008955	

Time taken: 8.824 seconds, Fetched: 8 row(s)

```
Time taken: 6.064 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((LateAircraftDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230301032238_71ce1c3f-c7c9-4051-8411-ee70af50a491
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.22 s

PK	year	_c1
0003	37.924225963706164	
0007	45.252432744291134	
0009	33.585314999939314	
0010	41.331052610239794	
0004	31.662176952308105	
0005	49.490838422749654	
0006	46.03070724001735	
0008	37.2255555400794	

Time taken: 7.259 seconds, Fetched: 8 row(s)

```

Time taken: 7.259 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((SecurityDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230301032327_7ff689cf-b10b-4133-86d6-af641029cde1
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)

```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

```

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 7.63 s
-----
PK
year      _c1
2003      0.0
2007      0.22865853658536586
2009      0.0
2010      0.0
2004      0.0
2005      0.0
2006      0.0
2008      0.0
Time taken: 8.462 seconds, Fetched: 8 row(s)
hive>

```

Year wise carrier delay from 2003-2010

```

hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230301032453_6c09b55a-3e54-49a4-9102-e3d3f291be33
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)

```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

```

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.98 s
-----
PK
year      _c1
2003      24.557549755575373
2007      19.850007017971283
2009      28.33058554239575
2010      21.89310246015957
2004      43.64459443230066
2005      28.01977637202288
2006      30.453296261292596
2008      28.88346981456985
Time taken: 7.641 seconds, Fetched: 8 row(s)
hive>

```

```

Time taken: 7.041 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230301032509_238e604d-3957-4524-b096-c8fe135d55c7
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)

```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

```

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 8.85 s
-----
PK
year      _c1
2003      24.557549755575373
2007      19.850007017971283
2009      28.33058554239575
2010      21.89310246015957
2004      43.64459443230066
2005      28.01977637202288
2006      30.453296261292596
2008      28.88346981456985
Time taken: 9.538 seconds, Fetched: 8 row(s)
hive>

```

```

Time taken: 9.338 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230301032528_66193f56-aded-497a-bd07-e5d8cfc49d42
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)

```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

```

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.03 s
-----
PK
year      _c1
2003      24.557549755575373
2007      19.850007017971283
2009      28.33058554239575
2010      21.89310246015957
2004      43.64459443230066
2005      28.01977637202288
2006      30.453296261292596
2008      28.88346981456985
Time taken: 6.66 seconds, Fetched: 8 row(s)
hive>

```



```

hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230301032545_8df5f5a3-eee0-47b0-b288-decb8890f40a
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)

```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

```

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 7.16 s
OK
year      _c1
2003      24.557549755575373
2007      19.850007017971283
2009      28.33058554239575
2010      21.89310246015957
2004      43.64459443230066
2005      28.01977637202288
2006      30.453296261292596
2008      28.88346981456985
Time taken: 7.946 seconds, Fetched: 8 row(s)
hive>

```

```

hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230301032600_d71dfc55-2e58-4f7d-a91f-e29cb4e22101
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1677639775259_0001)

```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

```

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.60 s
OK
year      _c1
2003      24.557549755575373
2007      19.850007017971283
2009      28.33058554239575
2010      21.89310246015957
2004      43.64459443230066
2005      28.01977637202288
2006      30.453296261292596
2008      28.88346981456985
Time taken: 7.336 seconds, Fetched: 8 row(s)
hive>

```

Year wise NAS delay from 2003-2010

```

hive> SELECT Year, avg((NASDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305162856_7c6c14cb-e0e8-4d64-a5fd-41b017dad83f
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

```

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 7.02 s
OK
year      _c1
2003      29.686276314267346
2007      30.625925917941924
2009      37.63093330628511
2010      33.87351363404217
2004      18.24570801709598
2005      16.63068805373129
2006      18.119312329937703
2008      30.16552562594132
Time taken: 7.97 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((NASDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;

```

```

hive> SELECT Year, avg((NASDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305162920_a047b5fb-02c6-4270-ba16-14ee9cc3df95
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

```

VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.85 s
OK
year      _c1
2003      29.686276314267346
2007      30.625925917941924
2009      37.63093330628511
2010      33.87351363404217
2004      18.24570801709598
2005      16.63068805373129
2006      18.119312329937703
2008      30.16552562594132
Time taken: 7.716 seconds, Fetched: 8 row(s)

```

```

Time taken: 7.718 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((NASDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305162938_90f4d64d-2a74-40f4-b90e-ed2c0fb9a3e9
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED   1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED   2         2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.73 s
-----
OK
year      _c1
2003      29.686276314267346
2007      30.625925917941924
2009      37.63093330628511
2010      33.87351363404217
2004      18.24570061769958
2005      16.63068805373129
2006      18.119312329937703
2008      30.16552562594132
Time taken: 7.404 seconds, Fetched: 8 row(s)
hive>

```

```

Time taken: 7.718 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((NASDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163044_8434d430-6c4a-4343-9f95-427d7a810ddb
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED   1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED   2         2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 5.90 s
-----
OK
year      _c1
2003      29.686276314267346
2007      30.625925917941924
2009      37.63093330628511
2010      33.87351363404217
2004      18.24570061769958
2005      16.63068805373129
2006      18.119312329937703
2008      30.16552562594132
Time taken: 7.203 seconds, Fetched: 8 row(s)
hive>

```

```

Time taken: 7.718 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((NASDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163106_00368720-14a4-4e64-b414-2d2056f340dc
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED   1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED   2         2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.64 s
-----
OK
year      _c1
2003      29.686276314267346
2007      30.625925917941924
2009      37.63093330628511
2010      33.87351363404217
2004      18.24570061769958
2005      16.63068805373129
2006      18.119312329937703
2008      30.16552562594132
Time taken: 7.662 seconds, Fetched: 8 row(s)
hive>

```

Year wise Weather delay from 2003-2010

```

Time taken: 6.836 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163224_6bf453e2-e537-4182-9e25-8c320cf7f2ef
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED   1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED   2         2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.23 s
-----
OK
year      _c1
2003      7.8319479664511205
2007      4.042975783210287
2009      0.45316615137982363
2010      2.902351295584664
2004      6.4475279976916555
2005      5.85869715149616
2006      4.588604183967953
2008      3.7254490054008955
Time taken: 6.836 seconds, Fetched: 8 row(s)
hive>

```

```
hive> SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_Flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163258_0de24ce1-325e-4ed8-a1ea-000ca73a68ea
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1        1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2        2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.99 s
-----
OK
year      _c1
2003      7.8319479664511205
2007      4.042975783210287
2009      0.45316615137982363
2010      2.9023312955584664
2004      6.4475279976916555
2005      5.85069715149616
2006      4.588604183967953
2008      3.7254490054008955
Time taken: 7.646 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_Flights WHERE Year is NOT NULL GROUP BY Year;
```

```
hive> SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_Flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163318_99b1bb87-d4cb-4ca1-b48c-7c862a3cb1d6
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1        1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2        2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.05 s
-----
OK
year      _c1
2003      7.8319479664511205
2007      4.042975783210287
2009      0.45316615137982363
2010      2.9023312955584664
2004      6.4475279976916555
2005      5.85069715149616
2006      4.588604183967953
2008      3.7254490054008955
Time taken: 6.674 seconds, Fetched: 8 row(s)
hive>
```

```
hive> SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_Flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163344_183b7dd2-e4f2-4f76-9374-4eec107248a6
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1        1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2        2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 7.24 s
-----
OK
year      _c1
2003      7.8319479664511205
2007      4.042975783210287
2009      0.45316615137982363
2010      2.9023312955584664
2004      6.4475279976916555
2005      5.85069715149616
2006      4.588604183967953
2008      3.7254490054008955
Time taken: 7.865 seconds, Fetched: 8 row(s)
hive>
```

```
hive> SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_Flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163439_cc0aa1b5-32e4-4be8-9481-282a42d85199
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1        1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2        2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.51 s
-----
OK
year      _c1
2003      7.8319479664511205
2007      4.042975783210287
2009      0.45316615137982363
2010      2.9023312955584664
2004      6.4475279976916555
2005      5.85069715149616
2006      4.588604183967953
2008      3.7254490054008955
Time taken: 7.11 seconds, Fetched: 8 row(s)
hive>
```

Year wise late aircraft delay from 2003-2010

```
hive> SELECT Year, avg((LateAircraftDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163652_f3f1527b-9539-4f25-ad8b-71d08faf5ff4
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1        1        0        0        0        0
Reducer 2 ..... container  SUCCEEDED    2        2        0        0        0        0
-----
VERTICES: 02/02  [=====] 100% ELAPSED TIME: 6.42 s
-----
OK
year      _c1
2003      37.924225963706164
2007      45.252432744291134
2009      33.585314999939314
2010      41.331052610239794
2004      31.662176952308105
2005      49.490838422749654
2006      46.838787224801735
2008      37.2255555408794
Time taken: 6.992 seconds, Fetched: 8 row(s)
hive>
```

```
hive> SELECT Year, avg((LateAircraftDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163712_66233be5-25f2-4e41-ad43-dfbdb0e8ced06
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1        1        0        0        0        0
Reducer 2 ..... container  SUCCEEDED    2        2        0        0        0        0
-----
VERTICES: 02/02  [=====] 100% ELAPSED TIME: 6.87 s
-----
OK
year      _c1
2003      37.924225963706164
2007      45.252432744291134
2009      33.585314999939314
2010      41.331052610239794
2004      31.662176952308105
2005      49.490838422749654
2006      46.838787224801735
2008      37.2255555408794
Time taken: 7.393 seconds, Fetched: 8 row(s)
hive>
```

```
hive> SELECT Year, avg((LateAircraftDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163734_43a4d71e-67a8-434b-b742-11fd694ba2a9
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1        1        0        0        0        0
Reducer 2 ..... container  SUCCEEDED    2        2        0        0        0        0
-----
VERTICES: 02/02  [=====] 100% ELAPSED TIME: 7.11 s
-----
OK
year      _c1
2003      37.924225963706164
2007      45.252432744291134
2009      33.585314999939314
2010      41.331052610239794
2004      31.662176952308105
2005      49.490838422749654
2006      46.838787224801735
2008      37.2255555408794
Time taken: 7.645 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((LateAircraftDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
```

```
hive> SELECT Year, avg((LateAircraftDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163752_e3a50a03-2a65-4912-b752-efdbafdd6d309
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1        1        0        0        0        0
Reducer 2 ..... container  SUCCEEDED    2        2        0        0        0        0
-----
VERTICES: 02/02  [=====] 100% ELAPSED TIME: 6.70 s
-----
OK
year      _c1
2003      37.924225963706164
2007      45.252432744291134
2009      33.585314999939314
2010      41.331052610239794
2004      31.662176952308105
2005      49.490838422749654
2006      46.838787224801735
2008      37.2255555408794
Time taken: 7.252 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((LateAircraftDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
```

```
hive> SELECT Year, avg((LateAircraftDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163819_6fb463b1-861c-4201-8bae-4c09190c3aef
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES   MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1        1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2         2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.47 s
-----
OK
year      _c1
2003     37.924225963706164
2007     45.252432744291134
2009     33.585314999939314
2010     41.331052610239794
2004     31.062176952308105
2005     49.490830422749654
2006     46.830787224801735
2008     37.2255555408794
Time taken: 7.036 seconds, Fetched: 8 row(s)
```

Year wise security delay from 2003-2010

```
hive> SELECT Year, avg((SecurityDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163849_c76e5687-7226-4c77-9a97-548d69d3eeb
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES   MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2         2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 7.14 s
-----
OK
year      _c1
2003      0.0
2007      0.22865853658536586
2009      0.0
2010      0.0
2004      0.0
2005      0.0
2006      0.0
2008      0.0
Time taken: 7.674 seconds, Fetched: 8 row(s)
hive>
```

```
hive> SELECT Year, avg((SecurityDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305163943_b864fdb-dcb8-444d-be91-fc2adbdd9c4
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES   MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2         2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.53 s
-----
OK
year      _c1
2003      0.0
2007      0.22865853658536586
2009      0.0
2010      0.0
2004      0.0
2005      0.0
2006      0.0
2008      0.0
Time taken: 7.129 seconds, Fetched: 8 row(s)
hive>
```

```
hive> SELECT Year, avg((SecurityDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305164009_05f0918c-dc78-4ddf-8c03-ad16150e56f4
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES   MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2         2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.87 s
-----
OK
year      _c1
2003      0.0
2007      0.22865853658536586
2009      0.0
2010      0.0
2004      0.0
2005      0.0
2006      0.0
2008      0.0
Time taken: 7.505 seconds, Fetched: 8 row(s)
hive>
```

```
Time taken: 7.771 seconds, Fetched: 8 row(s)
hive> SELECT Year, avg((SecurityDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305164028_1265275c-b960-400a-876c-dbf644e52bc6
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2         2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 7.04 s
-----
OK
year      _c1
2003      0.0
2007      0.22865853658536586
2009      0.0
2010      0.0
2004      0.0
2005      0.0
2006      0.0
2008      0.0
Time taken: 7.771 seconds, Fetched: 8 row(s)
hive>
```

```
hive> SELECT Year, avg((SecurityDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
Query ID = hadoop_20230305164116_e2069003-2715-4f04-b88a-e223dbb67912
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1678032641994_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2         2         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 6.13 s
-----
OK
year      _c1
2003      0.0
2007      0.22865853658536586
2009      0.0
2010      0.0
2004      0.0
2005      0.0
2006      0.0
2008      0.0
Time taken: 6.649 seconds, Fetched: 8 row(s)
hive>
```

8. Terminating the Cluster and cleaning the S3 bucket

The screenshot shows the AWS Management Console interface for an Amazon EMR cluster named 'HiveCluster'. The cluster is in a 'Terminating' state, indicated by a red 'Terminating' label and the text 'Terminated by user request'. The console displays various tabs for the cluster, including Summary, Application user interfaces, Monitoring, Hardware, Configurations, Events, Steps, and Bootstrap actions. The Summary tab is selected, showing details such as the cluster ID (j-1SG29RU1GHMO), creation date (2023-03-01 08:26 UTC+5:30), and the fact that it was terminated after the last step completed. It also lists the master public DNS, tags, and application user interfaces. The Configuration details section shows the release label (emr-5.36.0), Hadoop distribution (Amazon 2.10.1), and applications (Hive 2.3.9, Hue 4.10.0, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2). The Network and hardware section shows the availability zone (us-east-1d) and subnet (subnet-08cdd0b1e5aa15171). The Security and access section shows the key name (vockey) and EC2 instance profile (EMR_EC2_DefaultRole).

The screenshot shows the AWS Management Console interface for the Amazon S3 service. A green banner at the top indicates 'Successfully deleted bucket "hiveclusterassignmnet1"'. The console displays the 'Buckets' section, showing an 'Account snapshot' with metrics such as total storage (3.0 MB), object count (394), and average object size (7.7 KB). Below the account snapshot, there is a list of buckets. The bucket 'aws-logs-055500521311-us-east-1' is listed with its AWS Region (US East (N. Virginia) us-east-1), Access (Objects can be public), and Creation date (January 26, 2023, 18:11:58 UTC+05:30).

aws

Services

Search

[Alt+S]

Global

voclabs/user2377842=munzima.23@uom.lk @ 0555-0052-1311

Successfully emptied bucket "hiveclusterassignmnet1"
View details below. If you want to delete this bucket, use the [delete bucket configuration](#).

Empty bucket: status

CancelExit

The details below are no longer available after you navigate away from this page.

Summary

Source	Successfully deleted	Failed to delete
s3://hiveclusterassignmnet1	✔ 3 objects	0 objects

Failed to delete (0)

Find objects by name

< 1 >

Name	Prefix	Version ID	Type	Last modified	Size	Error
No failed object deletions						

FeedbackLanguage

© 2023, Amazon Web Services, Inc. or its affiliates. PrivacyTermsCookie preferences